

What is claimed is:

1. A method of predicting a scaffold of a protein comprising a query sequence,

wherein said method uses a database which contains environmental information on the side chain of each amino acid residue contained in the amino acid sequence of each reference protein whose three-dimensional structure is known or predictable, and wherein said method comprises the step of:

conducting matching based on the environmental information on each amino acid residue of each reference protein and hydrophobicity or hydrophilicity property of the side chain of each amino acid residue of the query sequence, and

choosing at least one template protein among the reference proteins that has high similarity in three-dimensional structure to the protein comprising the query sequence.

2. The method according to claim 1, wherein the amino acid sequence of each of the reference proteins is divided into two or more segment sequences comprising two or more continuous amino acid residues based on the three-dimensional structure of the reference protein.

3. The method according to claim 2, wherein the amino acid sequence of each of the reference proteins is divided into one or more core segment sequences which substantially participate in the formation of hydrophobic core, and into one or more sub segment sequences which do not substantially participate in the formation of hydrophobic core.

4. The method according to ~~any one of claims 1 to 3~~, wherein the matching is conducted based on the information on the degree of burial into the inside of the protein or the degree of exposure

to the protein surface of the side chain of each amino acid residue in the reference protein, together with the properties of hydrophobicity or hydrophilicity of each amino acid residue in the query sequence.

a 5. The method according to ~~any one of claim 1 to 4~~, wherein the matching is performed by sliding one or more core segment sequences of the reference protein on the query sequence without consideration of any gaps except those at one end or both ends of the core segment sequences.

6. The method according to claim 5, wherein the gap is a deletion or addition of one or more amino acid residues.

7. The method according to ~~any one of claims 1 to 6~~, wherein the matching comprises the steps of:

Sub B3 007130 09508827 2880560
(a) the step of matching by sliding one or more core segment sequences on the query sequence, optionally considering gaps at one end or both ends of the core segment sequences, provided when two or more core segment sequences are used, the core segment sequences are placed in the order of appearance on the amino acid sequence of the reference protein; and

(b) subsequent to the above step (a), the step of sliding one or more sub segment sequences on the query sequence, optionally considering one or more gaps.

a 8. The method according to ~~any one of claims 1 to 7~~, wherein the optimum matching is selected based on calculated scores obtained from the environmental information on the side chains of the amino acid residues of the reference protein and the hydrophobicity parameters of the corresponding amino acid residues on the query sequence.

9. The method according to claim 8, which further comprises the step of normalizing the above scores by using a self matching

score for the reference protein.

a 10. The method according to ~~any one of claims 1 to 9~~, which further comprises the step of constructing the three-dimensional structure of the protein comprising the query sequence.

Sub B4
a 11. A database which contains environmental information on the side chains of amino acid residues of one or more reference proteins with known or predictable three-dimensional structures, and which is used for the method according to ~~any one of claims 1 to 10~~.

09508527.071700